1991

Multi-Sample Confirmatory Factor Analysis of NBAS Data: An Analysis of the Stability of Latent Structures across Different Samples of Newborn Infants

Scott Daniel Azuma
Loyola University Chicago

MULTI-SAMPLE CONFIRMATORY FACTOR ANALYSIS OF NBAS DATA: AN ANALYSIS OF THE STABILITY OF LATENT STRUCTURES ACROSS DIFFERENT SAMPLES OF NEWBORN INFANTS

by

Scott Daniel Azuma

A Dissertation Submitted to the Faculty of the Graduate School of Loyola University of Chicago and the Erikson Institute in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

May

1991

ACKNOWLEDGEMENTS

The author wishes to thank his committee members, Dr. Jack Kavanagh, Dr. Frances Stott, Dr. Robert Halpern, and Dr. Dan Griffith for their guidance in the completion of the dissertation.

My most sincere gratitude is accorded to Dr. Kathleen Malee for her friendship and encouragement throughout my graduate career.

Finally, deepest gratitude to my wife Cathern Grisham for putting up with me during all of this.

This work is dedicated to my parents Steven Yoneo and Tosuko Azuma.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

Infancy

The study of infant behavior

Infancy generally refers to the period of life between birth and the emergence of language (approximately in the second year of life). This time period is characterized by rapid developmental and physical change. The study of infants therefore provides psychologists with a developmental arena within which general developmental principles can be observed.

The beginning of infant studies can be traced back to Tiedemann's baby biography Record of an Infant's Life (1787) (Lamb & Bornstein, 1987). This 18th century publication is considered one of the first scientific studies of infant behavior (Hay, 1986). Darwin's publication of A Biographical Sketch of an Infant (1877) was also influential in emphasizing the importance of infancy. Darwin's comprehensive outline of his infant's development made it clear that the study of infants was critical to the overall understanding of human behavior.

While the infant has long been a subject of fascination to psychologists, it is only recently that infant capabilities have been more fully understood. The slow growth of knowledge can be attributed, in part, to psychology's early and pervasive conceptualization of the infant as a simple creature. Many of the major discoveries concerning infant behavior have occurred since psychologists acknowledged the infant as a behavioral complex, active participant within development processes.

As psychologists increasingly believed that infants were active in shaping their own development, the drive to understand infant behavior increased. The attention directed towards infancy over the past 30 years has produced a wealth of studies which have identified various infant abilities, explored varying conceptualizations of infancy, and clarified general themes within this stage of development.

The conceptualization of infant development

Infancy is considered a microcosm of developmental psychology as a whole. Therefore, over the years, infant studies have reflected a wide range of theoretical viewpoints. Despite the lack of theoretical unanimity, certain themes have guided the developmental study of infants. The most common theme involves the contribution of environmental (nurture) and innate factors (nature) to human development.

Early developmental theorists tended to see development as either driven by environmental factors or resulting from an innate, maturational mechanism at place since birth. This type of nature-nurture debate continues even today although contemporary theorists generally account for both nature and nurture effects within their developmental models.

The direct nature or nurture model was originally challenged in the 1950's by Anne Anastasi (1959), who argued for an interactive model of environmental and innate effects. The interactive model was additive, hypothesizing that nature and nurture effects together shaped development. This model was highly influential in changing research questions to how environmental and innate forces interact rather than which of the two is primarily responsible for development (Lamb & Bornstein, 1987).

Developmental work by Arnold Sameroff and Micheal Chandler in the 1970's expanded the interaction model into a transactive model. This model proposed that constitutional and environmental effects influenced each other in a continual developmental process (Sameroff & Chandler, 1975). The effect of an experience depended upon the nature of the experience and the current status of the individual (Lamb & Bornstein, 1987).

The transactional model is currently widely accepted in developmental literature (Lamb & Bornstein, 1987). Use of this perspective has facilitated investigation of the entire spectrum of human development. The transactional model allows for the exploration of how human characteristics influence experiences and how experience simultaneously influences ultimate development (Lamb & Bornstein, 1987).

Today's conceptualization of the infant reflects a transactional model of development. Infants are thought to develop innate capacities within an environment which either facilitates or inhibits actions (Sameroff & Chandler, 1975). Because social and environmental factors are highlighted, parents and other members of the social matrix are considered critical components to the development of the infant.

The transactional model has also influenced psychology's concept of the neonate. The neonatal period is the earliest period of infancy encompassing the first 28 days of extrauterine life (Francis, Self, & Horowitz, 1987). Early conceptualizations of the neonate reflected a neurological model and essentially assessed the infant in terms of reflexive integrity. Transactional theory, on the other hand, focused attention upon neonatal behaviors as cues for and responses to the caretaker. These mutual influences were recognized as vital components of early development. In keeping with this view, neonatal assessments were gradually developed to examine the more complex and socially driven neonate. Over time, the neonatal period had been recognized as an important window to the understanding of infant behavior.

The publication of the Neonatal Behavioral Assessment Scale (NBAS) (Brazelton 1973, 1984) was an important contribution to the field of neonatal assessment. This behavioral assessment clearly reflected transactional theory as it considered the neonate as an organism which was actively involved in shaping and being shaped by his or her environment. The NBAS was the first neonatal examination to recognize that the individuality of the infant may influence the outcome of the relationship with his or her caretakers (Brazelton, 1973).

Since its original publication, the NBAS has proved to be valuable for highlighting the behavioral characteristics which neonates bring to the social situation. The NBAS provides a look backwards into the interuterine experience and a look forward to what the neonate brings to the caretaker and environment (Brazelton, Nugent, & Lester, 1987).

Overview of the investigation

The preceeding outline of the conceptualization of the neonate and infant was meant to introduce the topic of neonatal assessment with the NBAS. Assessments like the NBAS have contributed to and been influenced by the current conceptualization of the infant (Francis, Self, & Horowitz, 1987). Observations and measurement of the neonate continue to highlight information which is vital to the overall understanding of infant development.

Assessment of developmental constructs in infancy is difficult because of the multi-factorial nature of the entities examined. Therefore, infant researchers must constantly strive to show that the measurements of phenomena are both reliable and valid. This is especially critical in neonatal studies where the majority of behavioral variables are hypothetical entities.

Empirical definition of hypothetical constructs is the focus of the current investigation of the NBAS. Review of the literature will establish that the examination has been widely used and fits well with the current conceptualization of the infant. The review will also document that important psychometric issues such as construct validity and item reliability have not been adequately documented for the NBAS.

The investigation is based upon the premise that clarification of psychometric issues is necessary in order for the NBAS to be considered an appropriate assessment of infant behavior. The primary goal of this investigation is to empirically determine the reliability and validity of commonly used NBAS constructs (orientation, motor functioning, state range, state organization, and autonomic functioning). The information generated from the analysis is considered important documentation of the measurement capacity of the NBAS.

Quantifying measurements is necessary for reasons other than empirical knowledge. Many contemporary developmental risk situations involve infants and neonates. These situations include prenatal use of drugs and alcohol, pervasive poverty, prematurity, and low birthweight. Neonatal assessment with the NBAS has become a standard procedure for studies documenting these types of developmental risk factors.

Empirical documentation of psychometric qualities will improve psychology's understanding of both normal and at risk neonatal behavior. This information will in turn help psychologists meet the changing developmental needs of neonates and infants.

CHAPTER II

REVIEW OF THE LITERATURE

The review of the literature is divided into four major sections. The first section provides background information concerning the Neonatal Behavioral Assessment Scale (NBAS). This section includes a description of the examination, examples of its usage, and a description of data reduction strategies. The second section reviews the psychometric issues of reliability and validity. The third section provides information concerning the factor analytic procedure. This section describes both exploratory and LISREL approaches to factor analysis. The final section reviews the reliability and validity of the NBAS and includes a summary of the review of literature.

Neonatal Behavioral Assessment Scale

The Neonatal Behavioral Assessment Scale (NBAS) (Brazelton, 1984) is currently the most well-known examination of the newborn infant. The examination is based upon the theoretical concept that the primary developmental task of the newborn period is the stabilization of alert periods (Als, 1978). The emergence of this capability is seen as a product of the continous interaction and differentiation among the autonomic, motor, state, and attentional functioning systems

(Als, 1983). The autonomic system is typically defined by behaviors which reflect homeostatic functions. The motor system is reflected in the infant's postural tone and active movements. State functioning is commonly divided into two components: state range and state regulation. State range is represented by the states of consciousness available to the infant and the pattern in which transitions occur. State regulation consists of the observable strategies employed by the infant to maintain a balanced state of consciousness. Attention is characterized by the infant's ability to orient to inanimate and animate visual and auditory stimuli (Als, 1983; Lester, 1984).

The NBAS is designed to be an interactive process rather than an assessment of the infant's isolated responses. That is, the NBAS is a test of the infant's general capacity to respond to external manipulation in a social context rather than a formal neurological examination (Brazelton, 1984). The pattern and quality of responses are decisive in determining the normalcy of the infant's behavior (Als, Tronick, Lester, & Brazelton, 1977).

The items on the NBAS are divided into two parts: elicited reflexive items and behavioral items. The reflexive items are based upon responses originally described by Prechtl and Beintema (1968). Each of the 16 reflexive items are rated on a 1-3 scale; 1 is considered a hypoactive response, 2 is normal, and 3 is considered hyperactive. The behavioral items are rated on a 9 point scale. The midpoint for most of the behavioral items represents the expected performance for a normal, fullterm infant who is 3 days old (Brazelton, 1984). Scoring criteria for each of these items vary from item to item and are provided in the manual. The infant's score on each of the behavioral items is based upon his or her best response or performance.

Use of the NBAS

Since its original publication (Brazelton, 1973), the NBAS has been used extensively in a wide variety of neonatal studies. For example, the NBAS has been used to document cross cultural differences in infant behavior in Puerto Rican infants (Garcia-Coll, Sepkosi, & Lester, 1981), Greek infants (Brazelton, Tryphonopoulou, & Lester, 1979), and African infants (Dixon, Keefer, Tronick, & Brazelton, 1982). In each of these studies, infant behavior as measured by the NBAS was compared with groups of white, American infants.

The findings from these and other studies revealed variations in neonatal behavior across the different cultures. For example, the study of Puerto Rican infants revealed that these infants were more active and sociable than white American infants (Garcia-Coll, Sepkoski, & Lester, 1981). Infants of African origin were found to be more motorically mature when compared to white infants (Dixon, Keefer, Tronick, & Brazelton, 1982).

The NBAS has also been used to document the effects of maternal medication on the newborn infant. The short term effect on infants has been documented by many different studies (Brazelton, Nugent, & Lester, 1987). For example, Kuhnert, Harrison, Linn, and Kuhnert (1984) found that use of lidocaine was associated with subtle differences in newborn behavior when compared with infants whose mother were administered chloroprocaine. Aleksandrowiscz and Aleksandrowiscz (1974) found neonatal behavioral effects of maternal medication through the first month of life. Other studies have suggested a synergistic effect of maternal medication and stress factors on early infant behavior. Lester et al. (1982) for example, found that obstetric medication in combination with length of labor, parity, and ponderal index affected the behavior of infants as measured by the NBAS. Woodson and DaCosta (1980) found that length of labor, in combination with medication variables significantly affected irritability scores on the NBAS.

Although the NBAS was designed to assess full-term, healthy infants, it has also been used to examine the effects of risk factors such as prematurity and in utero exposure to drugs. For example, Scanlon, Scanlon, and Tronick (1984) used the used the NBAS to investigate the relationship between neonatal factors and preterm behavior. These authors found significant correlations between NBAS scores on the fourteenth day of life and gestational age and birthweight. Lester, Emory, Hoffman, and Eitzman (1976) used the NBAS in a multivariate study to determine the relationships among early behavior and a number of risk factors including low birthweight. In this study, low birthweight and mother's age were found to be significantly related to attentional/orientation ability. Ferrari, Grosoli, Fontane and Cavazutti (1983) used the NBAS to document the behavioral abilities of preterm infants who had reached 40 weeks conceptional age. The results of this study indicated that preterm infants showed weaker orientation, motor performance, regulation of state, and autonomic regulation than fullterm infants. Paludetto et al. (1984) also documented the qualitative behavioral differences in their study of preterm and fullterm infants.

Additional research with preterm infants has been based upon an adapted version of the NBAS developed by Als, Lester, Tronick, and Brazelton (1982). This expanded and revised version of the NBAS provides items specifically defined to document preterm infant behaviors. The examination has been used to quantify and qualify the behaviors of preterm infants (Als, 1983; Sell et al. 1980) and to develop intervention strategies for preterm infants in the Special Care Nursery (SCN) (Als, 1986).

The NBAS has frequently been used to document the effects of maternal substance abuse. Infants born to alchohol addicted mothers were studied by Streissguth, Martin and Barr

(1983) who noted that maternal alcohol use in midpregnacy was related to poor habituation and to significantly lower arousal. Chasnoff, Hatcher, and Burns (1982) used the NBAS to study the effects of 4 different types of maternal addiction. The findings indicated that all four groups had relatively poor state control as compared with a control group. Strauss et al. (1976) used the NBAS to determine that narcotic addicted infants had significantly decreased behavioral organization when compared to control infants. More recently, the NBAS has been used to document the effect of maternal cocaine use on early newborn behavior (Chasnoff, Griffith, MacGregor, Dirkes, & Burns 1989).

One of the difficulties in using the NBAS concerns the organization of the individual NBAS items for data analysis. The large number of items require some type of clustering in order to avoid multiple single item analyses. Two clustering models have been utilized with NBAS data; an a priori model and Lester's clusters.

The a priori system was derived by Adamson, Als, Tronick and Brazelton (1975). These authors hypothesized that NBAS items represent 4 processes: 1) interactive; 2) state organization; 3) physiological; and 4) motoric. Items are combined within each of these dimensions in a qualitative manner to produce a rating of worrisome, average, or optimal. The 4 dimension scores are then combined to produce a summary score that represents the overall behavior of the infant (Als et al. 1979). This 3 point ordinal system was expanded by Sostek and Anders (1977) to a 5 point system in order to increase the range of the scale and allow for more descriptive classification of infants.

The a priori clusters were criticized on the grounds that the groupings of items were not clinically understandable (Brazelton, Nugent, & Lester, 1987). In addition, it was felt that the 3 point system decreased the sensitivity of the NBAS items and therefore reduced the diagnostic nature of the examination (Jacobson et al., 1984).

The most widely used data reduction system for the NBAS was developed by Lester (1983). Lester hypothesized a 7 cluster system based upon previously reported factor analyses and his own factor studies and statistics (Brazelton et al., 1987). Within this system, all behavioral items are recoded so that a higher score represents optimal performance. An alertness dimension is represented by the orientation and the alertness items. Motor functioning is defined with the tonus, motor maturity, pull to sit, defensive reaction, and activity items. Range of state consists of the peak of excitement, rapidity of buildup, and irritability items. Cuddliness, consolability, self-quieting, and hand-to-mouth define the regulation of state system. Autonomic functioning is defined with the tremulousness, startles, and lability of skin color items (Lester, 1983). Within each of these behavioral clusters, items are averaged together to form a summary score.

Reflexive items represent their own separate cluster and are summarized by counting the number of abnormal reflexes.

In contrast to the wide variety of studies which document the usage of the exam, little information is available concerning the validity or the reliablity of the NBAS. Documentation of these issues are critical for the usage of an evaluation as a research mechanism.

Reliability and Validity

Reliability is generally defined in terms of how well measurements can be reproduced. Measurements are intended to be stable over a variety of conditions. If a measurement approach provides the same result regardless of variations in the situation it is considered a reliable measurement (Nunnally, 1978).

Estimates of reliability are generally based upon the average correlation among the items on the test. Coefficient Alpha (Cronbach, 1951) is the traditional means for determining this average internal consistency. Coefficient Alpha provides a good estimate of reliablity in most test cases because the major source of measurement error is associated with content (Cronbach, 1951). If Alpha is low then it can be assumed that the items on the test have little in common and therefore are suspect in terms of content representation (Anastasi, 1982).

In addition to internal consistency, tests should be evaluated in terms of test-retest reliability. Correlations between test items taken at different times should be high, given that the attribute being measured is known to be relatively stable. It should be noted that documentation of test-retest consistency is not as critical as internal consistency. A reliable test is based upon items which correlate with one another. If Coefficient Alpha is low for a test, a relatively high correlation between tests does not constitute test reliability (Nunnally, 1978).

Validity is commonly defined in terms of how well the instrument does what it intended to do (Nunnally, 1978). Validation of an instrument is generally subdivided into three distinct but related categories: 1) predictive validity, 2) content validity, and 3) construct validity. These three types of validity represent different issues concerning scientific generalization of the measurement. For example, predictive validity should be considered when the purpose of the instrument is the estimation of a behavior that is external to the measuring instrument itself. Content validity is important when determination of the adequacy of the domain of behavior is critical. Construct validity is necessary when the issue is empirical verification of a measure representing some aspect of behavior (Nunnally, 1978; Anastasi,1982).

Predictive validity concerns the extent to which one can generalize from scores on one variable (instrument) to scores on another variable (criterion). The correlation between the predictor test and the criterion variable denotes the degree of validity. A predictor test cannot be valid unless it has a significant correlation with the criterion variable. Although predictive validity is conceptually simple, it is frequently complicated by the lack of a appropriate criterion measure. In many situations with psychological measures, either no criterion exists or the criterion itself is not well defined (Anastasi, 1982).

In other situations, the validity of an instrument depends primarily on the adequacy with which a specified domain of content is sampled. A good example of this situation is a final examination for a school course. Obviously, the measurement cannot be validated in terms of predictive validity because the purpose of the test was not prediction. Validity is assessed in terms of how well the items chosen represented the content of the course.

In order for the items to be representative, the domain of study must be clearly specified (Anastasi, 1982). Clearly defined domains increase the likelihood of adequate item selection. In addition, content validity is established by indices of internal consistency. Moderate internal consistency among the items is considered necessary in order for the test to be valid (Nunnally, 1978). Moderate correlations with tests which measure similar domains are considered additional support for content validity. These types of correlations are not, however, the major component of content validity. Content validity is mainly determined heuristically rather than by a specific statistic (Nunnally, 1978).

The most important component of validity involves the construct (Angoff, 1988). Hypothetical variables must be adequately measured in order that they have interpretable meaning. Predictive and content validity support the measurement of a construct but are not themselves sufficient evidence for construct validation (Nunnally, 1978; Messick, 1988).

Construct validation begins with stating the meaning of the word or words representing the construct (ie what is "anxiety"). Once the domain is defined, the adequacy of the domain is assessed in terms of how well the observable measures correlate. If all the measures correlate highly, they most likely represent the same behavior or construct. If the measures tend to split into different sets of variables, a number of different constructs are being measured by the observable variables (Nunnally, 1978).

Evidence which shows that the construct behaves as expected is further indication of construct validity. How a construct behaves is determined in part, by the underlying theory with which the construct was derived. For example, a construct of sensorimotor ability theoretically should be related to later cognitive constructs. In order to validate the sensorimotor construct, it should be investigated in terms of its relationship to later cognitive constructs.

Factor analysis is one statistical procedure that can be used to provide evidence for each type of measurement validity. As a procedure, it provides empirical information regarding both the internal statistical structure of the instrument and the relationships these structures have with other constructs. Factor analytic procedures are therefore considered a major component in establishing measurement validity of an instrument (Nunnally, 1978).

Factor Analysis

History

The original conceptualization of factor analysis is generally attributed to Charles Spearman (1904, 1927) who developed a factor theory in conjunction with his study of intelligence. Spearman's factor model postulated that 2 factors represent human ability: a general factor (g) and a separate, unique factor (Tatsouka, 1988).

Continued study of factor theory and intelligence eventually concluded that Spearman's 2 factor model was too restrictive. Alternative theories were subsequently proposed and tested by Holzinger (1944, 1949) and Thurstone (1931, 1945, 1947) among others. Holzinger hypothesized a general factor and group factors (instead of specific factors) which each affected a small group of related abilities instead of just a single ability (Holzinger & Swinefield, 1937). Thurstone eliminated the general factor altogether and retained only the group factors, establishing a multiple factor or common factor theory of human abilites (Thurstone, 1945). Thurstone's theory did allow for the existence of unique factors; however, each of the unique factors were hypothesized to affected only a single ability.

Thurstone and his followers were highly influential in the development of the multiple factor procedure (Mulaik, 1986). Thurstone (1944) substantiated his multiple factor theory with factor studies which predicted the primary factors of intelligence. Thurstone's work was also influential in terms of rotation of solutions in that he felt that factors should be rotated to as simple a structure as possible in order that "correct factors" be identified (Mulaik, 1986).

Over the past thirty years, advances in computer technology have eased the computational burden of factor analysis making it an accessible statistical procedure. Recent theoretical advances have mainly involved development of different types of factor solutions such as the maximum likelihood etimation. Ongoing statistical work with these types and other types of solutions have made factor analysis an integral part of psychological and educational investigations (Harmon, 1967).

Exploratory Factor Analysis

Exploratory factor analysis conceptualizes latent factors in the following manner. For a given set of responses, X_1, X_2, X_3,....X_q one wishes to determine a set of underlying latent factors η_1, η_2 , η_3,...η_n, fewer in number than the observed variables. These latent factors hypothetically account for the intercorrelations of the observed variables such that no correlation would remain among the observed variables if the factors were partialled out (Kim & Mueller, 1978).

In factor analysis, each of the factors is correlated with one another and each of the factors have causal effects on the observed variables. Factors are considered common since their effects are shared in common with more that one of the observed variables. Unique factors are error variables; their effects are unique to one observed variable only. Unique factors are independent of the common factors and uncorrelated with one another (Harmon, 1967).

The main objective in an exploratory factor procedure is to determine the minimum number of common factors that would adequately reproduce the correlation among the observed variables. Statistically this involves applying some type of mathematical criterion which estimates the discrepancy between the assumed factor model and the actual data (Harmon, 1967). If the criterion determines that the discrepancy is minimal, the extraction of factors is stopped. If the discrepancy is not judged to be minimized, another factor is added and the process continues (Kim & Mueller, 1978).

There are many types of criteria for establishing the minimum number of factors. These methods include the maximum likelihood procedure, principal axis factoring, principle components analysis, alpha factoring, and image analysis

(Harmon, 1967). Historically, principle axis factoring has been the most commonly used extraction procedure in exploratory factor analysis (Harmon, 1967). Principle axis factoring is based upon the characteristic equation. Solving this equation produces eigenvalues and eigenvectors associated with the given correlation matrix (Tatsouka, 1988).

In the principle axis procedure, the correlation matrix is reduced; that is the ones in the diagonal are replaced by communalities (Tatsouka, 1988). Commonly used estimates of communalities are the squared multiple correlation for each variable with the remainder of the variables in the set, or the highest absolute correlation in a row of the correlation matrix (Kim & Mueller, 1978). After inserting these communalities in the main diagonal of the correlation matrix, factors are determined using the equation:

$$\det (R_1 - \lambda I) = 0$$

where R_1 is the correlation matrix with communality estimates in the main diagonal (Tatsouka, 1988).

The initial factoring step determines the minimum number of factors that can adequately account for the observed correlations. In these initial solutions, factors are orthogonal and arranged in descending order of importance. These properties are arbitrary impositions placed on the data in order that solutions are unique and defineable. The consequences of these restriction are that: 1) variables will have substantial loadings on more than one factor (factorial complexity) and 2) except for the first factor, the factors will be bipolar; some variables will have positive loadings on a factor while some will have negative (Kim & Mueller, 1978). In general, these restrictions make interpretation of the factors difficult. Therefore, the next step in the process involves finding simpler, more easily interpreted factors. This typically is accomplished through different types of rotations of the initial factor matrix.

The most common approach to rotation is to rely on some analytic rotation method that satisfies a particular criterion of simplicity. There are two subgroups within the analytic approach: orthogonal and oblique. In orthogonal rotations, the axes remain at 90° to each other. Oblique rotations allow the axes angles to vary. Different rotational methods are available within each of the subgroups of rotations.

An oblique solution is more general than an orthogonal rotation in that it does not arbitrarily impose the restriction that factors be uncorrelated. However, by introducing correlations among factors, the oblique solution can be difficult to interpret. Different oblique solutions are based upon both reference axis projections and primary pattern matrix formulas (Tatsouka, 1988).

Confirmatory Factor Analysis

When discussing factor analysis, it is important to distinguish between exploratory and confirmatory factor procedures. Exploratory procedures do not impose a specific model on to the data. These procedures therefore are used to detect relationships among variables and generate hypotheses. In a confirmatory procedure, a model is assumed apriori based upon previous knowledge concerning the variables. The objective then is to estimate the structure based upon the given model and statistically compare the generated structure to the actual data. If the structure is not found to vary significantly from the actual data, the model can be "confirmed" as a likely representation of the underlying structure for the given variables (Bentler, 1980).

Many of the advances in the area of confirmatory factor analysis have been attributed to the work of Karl Jöreskog (Bentler, 1980). Statistically, confirmatory factor analysis requires the estimation of the parameters of the model and the determination of the goodness of fit of the model. In general, it is impossible to estimate model parameters without a computer program because there is no algebraic solution and therefore iterative approximations that refine user-provided initial solutions are utilized (Bentler, 1980). Jöreskog provided the first practical computer solution to this statistical problem with his LISREL procedure (Jöreskog, 1973). Continued refinements of the program have made it the most widely used system for the analysis of latent variable and structural models (Biddle & Marlin, 1987).

The LISREL Procedure

The formal LISREL model can be delineated as follows. It is assumed that all variables are measured in deviations from their means. Random vectors $\eta' = (\eta_1, \eta_2, \eta_3, \ldots \eta_m)$ and $\xi' = (\xi_1, \xi_2, \xi_3, \ldots \xi_n)$ of latent dependent and independent variables, respectively are considered in the following system of linear structural relations:

$$\eta = B\eta + \Gamma\xi + \zeta,$$

where $B(m \times m)$ and $\Gamma(m \times n)$ are coefficient matrices and $\zeta' = (\zeta_1, \zeta_2, \ldots \zeta_n)$ is a random vector of residuals. The elements of B represent the direct effects of the η variables on other η variables. The elements of Γ represent the direct effects of ξ variables on η variables. It is assumed that ζ (error) is uncorrelated with ξ (independent variables) and that I-B is non-singular (Jöreskog & Sörbom, 1988).

The vectors η (dependent) and ξ (independent) are not observed, but instead vectors $y' = (y_1, y_2, y_3 \ldots y_p)$ and $x' = (x_1, x_2, x_3 \ldots x_q)$ are observed, such that

$$y = \Lambda_y\eta + \epsilon,$$

$$x = \Lambda_x\xi + \delta,$$

where ϵ and δ are vectors of error terms. These equations represent the regression of y on η and x on ξ. Thus the observed variables are refered to as x and y variables and latent dependent and independent variables as η and ξ respectively Jöreskog & Sörbom, 1988).

Estimation of the LISREL model

The term covariance matrix in the LISREL procedure refers to a general matrix which may be a matrix of moments about zero, a matrix of variances and covariances, or a matrix of correlations (Jöreskog & Sörbom, 1988). The procedure uses one of several kinds of estimation which include maximum likelihood (ML), unweighted least squares (ULS), instrumental variables (IV), two-stage least squares (TSLS), generalized least squares (GLS). The TSLS and IV methods are both non-iterative while the ULS and ML are obtained by an iterative procedure which minimizes a particular fit function by successively improving the parameter estimates. Initial values for the iterative procedures are provided by either the IV or the TSLS procedures (Hagglund, 1982; Jörsekog, 1983). These starting values are necessary for all iterative procedures in LISREL.

The ML procedure is derived from the maximum likelihood principle based on the assumption that the observed variables have a multinormal distribution (Jöreskog, 1967). The LISREL ML procedure minimizes the function:

$$F = \log||\Sigma|| + tr(S\Sigma^{-1}) - \log||S|| - (p + q).$$

The ULS procedure minimizes the function:

$$F=1/2 \ tr[(S-\Sigma)^2].$$

<u>LISREL Assessment of Model Fit</u>

The LISREL procedure provides several powerful tools for the assessment of fit of the hypothesized models and the actual data. The model assessment process involves analysis of the parameter estimates, the standard errors, squared multiple correlations, coefficients of determination, and the correlation among the parameter estimates.

Parameter estimates are evaluated in terms of size and direction (negative or positive). Standard errors are estimates of the precision of the parameter estimates and therefore should be small. LISREL T-values are also included in order to facilitate the analysis of the error terms. A T-value in LISREL is defined as the ratio of the parameter estimate and its standard error (Jöreskog & Sörbom, 1988). Examination of the correlation among the parameter estimates helps determine if one or more of the parameters are highly correlated. This type of correlation is an indication that the model is nearly not identified and that some of the parameters cannot be determined from the data. Squared multiple correlations are a measure of strength of a linear relationship of the observed variable to the underlying latent trait. The coefficient of determination is a measure of the strength of several relationships jointly (Jöreskog & Sörbom, 1988).

The second part of the evaluation of the model concerns the assessment of the overall fit of the model to the data. Within the LISREL procedure, fit is assessed by four separate measures: chi-square, goodness of fit (GFI), adjusted goodness of fit (AGFI), and the root mean square residual (RMSR).

The GFI is a measure of the amount of variance accounted for by the hypothesized model. The AGFI is the GFI adjusted for degrees of freedom in the model. The P-value associated with the reported chi-square is the probability of obtaining a chi-square larger than the value actually obtained given that the model is correct. The root mean square residual is a measure of the average of the fitted residual. This value is interpreted in relation to the sizes of the observed variances and covariances in the original matrix. (Jöreskog & Sörbom, 1988).

Examples of the use of LISREL

The LISREL procedure has frequently been used to validate constructs of behaviors. For example, McGaw and Jöreskog (1971) used the procedure to assess factorial invariance of scores on 12 aptitude and achievement tests across high and low socioeconomic groups. Newton, Komaeoka, Hoelter, & Tanaka-Matsumi (1984) examined the construct generality across males and females of self-report measures of anxiety with the LISREL procedure. In both of these studies, the LISREL simultaneous factor analysis procedure determined that the factors did not significantly vary across subgroups of samples. Other examples of multi-sample construct validation using the LISREL procedure are given by Sörbom (1974), McGaw and Jöreskog (1971), Werts et al., (1977) and by Fleishman and Beison (1987). Single sample construct validation studies using the LISREL procedure have been reported for attitude

(Bagozzi, 1978, 1981), self-concept (Marsh & Holevar, 1985), self-esteem (Maruyama et al., 1981), and psychological distress (Tanaka & Huba, 1984).

LISREL is most frequently utilized as a procedure for testing structural equational models. Examples of uses of the LISREL procedure for causal modeling include studies of academic expectations and school attainment (Entwisle & Hayduk, 1981), structural models on mathematics achievement (Ethington & Wolfe, 1986), and structural models for adolescent drug usage (Huba, Winegard, & Bentler, 1981).

Jöreskog's work has been influential in terms of the study of latent variables. Although Jöreskog's LISREL program is widely used, other computer model formulations for analysis of latent variables and structural equations are available and are considered by McDonald (1978), Bentler and Weeks (1980), and McArdle and McDonald (1984).

Validity and Reliability of the NBAS

(Studies which have explored underlying constructs of infant behavior using the NBAS have exclusively employed exploratory factor procedures.) For example, Sameroff, Krafchuck, and Bakow (1978) performed a multiple sample factor analysis on 300 infants from the Rochester Longitudinal Study. Four samples were derived using predominant state as the grouping variable. The factor analysis for each of the samples extracted 5 factors, of which the first 2 were most consistent across each of the samples. The first factor was an orientation structure which consisted of the orientation items and alertness. The second factor (labeled arousal) consisted of the peak of excitement, irritability, rapidity-of-buildup, and activity items. Factor 3 consisted of the motor maturity, tremulousness and lability-of-state items. Factor 4 consisted of the muscle tonus, pull-to-sit, and activity items. Factor 5 consisted of the self-quieting and hand-to-mouth items.

Kaye (1978) also reported a 5 factor solution in his study of 50 normal newborns. Orientation was found to be the first factor, consisting of the orientation and alertness items. Factor 2 consisted of the peak of excitement, rapidity-of-buildup, irritability, and lability-of-state items. Factor 3 consisted of the consolability, self-quieting and the hand-to-mouth items. Factor 4 consisted of the tonus, motor maturity, pull-to-sit, cuddliness, and activity items. The fifth factor in this analysis was not easily interpretable.

Osofsky's and O'Connell's (1977) factor analysis was performed with data collected on 328 normal infants. This factor analysis extracted 6 factors of which 3 were easily interpretable. Again, the first factor consisted of the orientation items. The second factor consisted of the peak of excitement, rapidity-of-buildup, irritability, and lability of-state items. The third factor was a habituation factor with 3 of the 4 habituation items loading on this factor.

Strauss and Rourke (1978) performed a factor analysis on a pooled sample of normal and at-risk infants (n=932). The results of this analysis were similar to those of previously reported factor studies. The analysis extracted 8 factors of which 3 were interpretable. Of interest, the first factor (orientation) only consisted of the visual items and alertness. The second factor consisted of the tonus, peak of excitement, irritability, and activity items. Habituation items comprised the third interpretable factor.

Lester et al. (1976) performed a factor analysis upon a sample of 52 infants who varied in term of perinatal risk factors. This analysis extracted 9 factors of which 2 accounted for the majority of the total variance. The first factor consisted of the orientation, alertness, tonus, and defensive reaction items. The second factor consisted of the consolability, peak of excitement, rapidity-of-buildup, irritability, lability-of-skin and state, and self-quieting items.

Streissguth et al. (1983) performed a factor analysis on 417 alcohol-exposed infants as part of a longitudinal study. This analysis extracted 6 factors with orientation loading on the first factor. Rapidity-of-buildup, lability-of-state, peak of excitement, and smiles made up the second factor. Factor 3 consisted of the 3 of the 4 habituation items.

Few factor studies have been reported on preterm infants. Sostek's and Anders' (1977) study of preterm infants revealed a factor structure similar to the those reported on fullterm infants. The first factor was comprised of the orientation items, with the second factor consisting of the consolability, self-quieting, and peak of excitement items. The third factor consisted of the habituation items.

(The findings of these studies suggest the existence of a 2 or 3 factor structure underlying newborn infant behavior. While item relationships varied across studies, the findings appear to substantiate orientation, arousal (state), and habituation dimensions. Of these 3 factors, orientation is the most consistent factor and generally accounts for the greatest amount of the total item variance. Beyond 2 or 3 factors, constructs of neonatal behavior do not appear to be clearly defined.

Additional information concerning the validity of the NBAS is currently limited. Predictive validity has been shown for the NBAS but is not consistent. For example, significant relationships between NBAS scores and later IQ were found in a population of preterm infants (Scarr-Salapatek & Williams, 1973). Lester (1983) reported that a sequence of NBAS scores were significantly related to 18 Month Bayley scores. In another study, Osofsky and Danzger (1974) reported measures of social responsiveness were significantly related to NBAS scores. Crockenberg (1981), noted that high neonatal irritabilty was found to be related to slower maternal responsiveness at 3 months Crockenberg, 1981). Waters, Vaughn, and Egeland (1980) found that low scores on the orientation, motor maturity, and regulation items predicted anxious and resistant attachment patterns at 1 year of age.

It should be noted that although significant predictive findings have been reported, the amount of variance accounted for by the NBAS is generally small (Horowitz & Linn, 1984). It is currently unclear how well the NBAS predicts and what types of later behaviors the assessment is related to.

In terms of reliability, the NBAS is generally reported to have high interrater reliability, provided that the examiners have themselves been trained appropriately (Francis, Self, & Horowitz, 1987; Nugent, Sepkoski, 1984). However, moderate to low test-retest reliability have been reported for the majority of NBAS items (Sameroff, Krafchuk, Bakow, 1978; Kesterman, 1981). Of note, many authors have attributed this test-retest fluctuation to normal variability within early infant behavior rather than a psychometric flaw within the NBAS itself (Horowitz, Sullivan, & Linn, 1978; Kaye, 1978; Lester, 1984). The lack of stability between examinations is seen as documentation of adaptive infant behavior and thus is indication of true rather than error variance (Lester, 1984).

Little is known about the reliability of Lester's clusters (Brazelton et al. 1987). Currently, only one article reports reliability information concerning this clustering system. These authors reported that internal consistency of the clusters, as measured by coefficient Alpha, ranged from

.81 (orientation) to .19 (autonomic stability). Test-retest reliabilities of the clusters ranged from .05 (reflexes) to .54 (autonomic stability) (Jacobson et al. 1984).

Summary of the Literature

The literature indicates that the NBAS can provide the researcher with a systematic way of eliciting behaviors which reflect the infant's current level of behavioral functioning. These types of measures provide key information concerning the infant's developing behavioral repetoire (Als, 1978). Currently, however, little information is available concerning the reliability or validity of the examination. Further information documenting the validity of the NBAS items and clusters are necessary in order for the examination to be considered a valid measurement of infant behavior. Specifically, if NBAS data is to be conceptualized and analyzed in terms of clusters such as defined by Lester, these constructs must be empirically defined. Review of the current literature does not confirm the constructs described by Lester. If these constructs do exist, it must be determined how well the NBAS items repesent them in order for the examination results to be interpreted correctly.

The review indicates that confirmatory factor analysis is an appropriate method for determining the reliability and validity of underlying, hypothetical constructs. The procedure allows the researcher to hypothesize a model and test it with statistics against the actual data. Currently, no studies are available which utilize this procedure with NBAS data. Information from the confirmatory factor analysis can therefore provide important evidence concerning the reliability and construct validity of NBAS. In terms of current usage of NBAS data, results from a confirmatory procedure can provide empirical evidence supporting Lester's model of factors or it will determine if alternative models of underlying constructs are more likely possibilities.

CHAPTER III

METHODOLOGY

The methodology chapter is divided into four sections. The first states the purpose of the study and is followed by the a statement of the study hypotheses. The third section describes the subjects involved in the current study. The final section summarizes the procedures involved in the study and includes a description of the construct models tested in the investigation.

Purpose

The theory underlying the NBAS implies that orientation, motor functioning, state range, state organization, autonomic functioning and habituation are underlying constructs of infant behavior. However, empirical confirmation of these hypothetical constructs remains lacking. The primary goal of this investigation was to establish a construct model of infant behavior based upon NBAS items. Specifically this involved using confirmatory factor analytic procedures to identify a construct model which adequately represented the underlying structure of infant behavior. Model adequacy was defined in terms of construct invariance across the different samples of infants. The determination of construct invariance is an important prerequisite for valid population comparisons

(Newton, et al., 1984) and is considered to be a critical component of measurement validity. Overall the information from this study provided important empirical evidence concerning the reliability and validity of the constructs underlying the NBAS. Information of this type was considered necessary in order that the measurement validity of the NBAS be established.

Hypotheses

Specific hypotheses were related to constructs of infant behavior as measured and defined by the NBAS. They included:

1. Lester's model of data reduction cannot be confirmed by multi-sample confirmatory factor analysis. The procedure does not find the factors defined by Lester (orientation, motor, state range, state regulation and autonomic functioning) to be consistently defined across different samples of infants.
2. A two factor model consisting of orientation items and state range items can be confirmed with the multi-sample confirmatory procedure. Confirmation of the model relied upon establishing construct invariance. Specifically, invariance was defined as follows:
   a. The number of factor constructs does not vary across populations.
   b. Parameter estimates of the observed variables which define the major constructs does not significantly vary across populations of infants.

c. The error associated with the observed variables does not significantly vary across populations of infants.

d. The correlation among the factor constructs does not significantly vary across populations of infants.

Note that factor invariance was tested incrementally. Each of the tests for invariance was based upon the confirmation of the previous step. For example, if the parameter estimates of the constructs were not found to be invariant across samples, the error and construct correlations were also assumed to vary significantly across samples.

Subjects

The subjects included samples from three different populations of newborn infants: 1) normal fullterm infants; 2) preterm infants; and, 3) fullterm infants who were exposed to drugs in utero. The preterm infants were sampled from infants who were admitted to the Special Care Nursery of Northwestern Memorial Hospital/Prentice Pavilion (n=166). Mean birthweight, gestational age, chronological age, and conceptional age at the time of testing are presented in Table 1. Percentages of the incidence of Respiratory Distress Syndrome (RDS) and Intraventricular Hemorrhage (IVH) for this sample are presented in Table 2. Grades of IVH were identified according to the criteria of Papile, Munsick-Bruno, and Schaefer (1983). Degrees of RDS were identified according to clinical and radiologic findings as defined in Table 2. All infants in the preterm sample were part of a comprehensive

TABLE 1

PERINATAL CHARACTERISTICS OF PRETERM SAMPLE (n=166)

	Mean	SD	Range
Gestational age (weeks)	29.5	2.4	25-37
Birthweight (grams)	1097.3	230.2	580-1500
Chronological age (weeks)	8.7	3.5	2-18
Conceptional age (weeks)	38.1	2.5	34-45

TABLE 2

INCIDENCE OF MEDICAL COMPLICATIONS

Intraventricular Hemorrhage (IVH)

Degree of Hemorrhage[a]	N	(percentage)
None	95	(57%)
Grade I	53	(32%)
Grade II	10	(6%)
Grade III	3	(2%)
Grade IV	5	(3%)

Respiratory Distress Syndrome (RDS)

Degree of RDS[b]	N	(percentage)
None	14	(8%)
Mild	40	(24%)
Moderate	52	(31%)
Severe	60	(37%)

[a] IVH graded according to Papile, Munsick-Bruno, & Schaefer (1983) criteria

[b] RDS criteria
Mild: less than 24 hours of mechanical ventilation.
Moderate: 1-5 days of mechanical ventilation.
Severe: more than 5 days of mechanical ventilation.

developmental follow-up program. Infants meeting the follow-up criteria received an NBAS examination prior to their discharge from the nursery as well as frequent outpatient developmental evaluations through the first 5 years of life.

The infants comprising the drug exposed sample were sampled from a population of infants enrolled in the Perinatal Center for Chemical Dependency (PCCD) program at Northwestern Memorial Hospital/Prentice Women's Hospital (n=267). Mothers enrolled in this project were followed prenatally and postnatally because of known chemical dependency. As part of the developmental follow-up, infants in this program received an NBAS examination prior to discharge from the newborn nursery. Mean birthweight and racial breakdown for these infants are reported in Table 3. Breakdown of the drug history of these infants is presented in Table 4.

Normal newborn data was collected from a sample of normal infants born at Beth Israel Hospital in Boston, Massachusetts (n=110). All infants were 40 weeks gestation at birth and were of normal birthweight. These infants were part of a control group for a cross-cultural study of newborn infants in Ireland.

TABLE 3
CHARACTERISTICS OF THE DRUG EXPOSED SAMPLE

Mean Birthweight (grams)	3700
Race (percentages)	
White	15%
Black	65%
Hispanic	15%
Other	5%

TABLE 4
BREAKDOWN OF DRUG HISTORY

Drug Used	Percentage of Population
Cocaine	15%
Marijuana	10%
Alcohol	5%
Cocaine/Marijuana/Alcohol	37%
Cocaine/Marijuana	17%
Cocaine/Alcohol	8%
Marijuana/Alcohol	8%

Procedures

All NBAS examinations were performed and scored by a certified Brazelton examiner. Each of the examiners for the preterm and drug exposed samples were familar with their respective populations. Of note, items added in the second edition of the NBAS (inanimate auditory and visual orientation and the 9 supplementary items) were not included in the investigation as some of the infants were tested prior to the publication of this edition.

For all analyses, NBAS items were rescaled such that higher scores represent more optimal functioning. Items which required rescaling were recoded in accordance with Lester (1983), and are presented in Table 5. In addition, for all analyses, habituation and reflex items were excluded. Habituation and reflex items are thought to represent cohesive factors and therefore were eliminated to reduce the number of variables in the analyses.

The remaining 22 NBAS items were analyzed in a confirmatory factor analysis using the LISREL VI procedure and a maximum likelihood solution (Jöreskog & Sörbom, 1988). Lester's model, the 2 factor model (orientation and state range) and four other alternative models were tested with the confirmatory procedure. These 4 other models were hypothesized a priori based in part upon the results of previous exploratory factor analyses. The models were categorized and labeled by the number of factors hypothesized.

TABLE 5
LESTER MODEL RESCALING[a]

Item	Rescale
Tonus	9/1=1; 8/2=2; 7/3=3; 4=4; 5=5; 6=6
Activity	9/1=1; 8/2=2; 7/3=3; 4/6=4; 5=5
Peak of Excitement	9/1=1; 8/2=2; 4/3=3; 7/5=4; 6=5
Rapidity of Buildup	9/1=1; 8/2=2; 7/3=3; 4=4; 5=5; 6=6
Irritability	9/1=1; 8=2; 7=3; 6=4; 5=5; 2,3,4=6
Lability of State	1,7,8,9=1; 5,6=2; 4=3; 3=4; 2=5
Tremors	Invert 9=1 (1=9) scale
Startles	Omit 1 then invert scale 2=9 (9=2)
Skin	9,1=1; 8=2; 7=3; 6=4; 5=5; 3,4=6; 2=7

[a] Lester, 1984

A summary of Lesters' model is presented in Table 6. A summary of the alternative models is presented in Table 7. Note that the hypothesized two factor model (orientation and state range) was designated Model 2B.

For each confirmatory analysis, covariance matrices were used as input data. Sample covariance were generated from the raw data using a SPSSX procedure (SPSS-X, 1987). Scales for each of the underlying factors were established by setting the value of 1 to the first variable on each factor.

The LISREL VI submodel for confirmatory factor procedures was a measurement model of x variables. There were no structural equations in this model and only one of the 3 LISREL equations are used in the computation. The model was denoted as follows:

$$x = \Lambda_x \xi + \delta$$

The assumptions of the model were that the ξ's and the δ's are random variables with zero means and that the δ's are uncorrelated with the ξ's. The model equation represents the regression of x on ξ such that the element λ_{ij} of Λ_x is the partial regression coefficient of ξ_j in the regression of x_i on ξ_1, ξ_2,.....ξ_n (Jöreskog & Sörbom, 1988).

TABLE 6
LESTER MODEL (FIVE FACTORS)

Orientation	Motor
Inanimate visual	Tonus
Inanimate auditory	Motor maturity
Animate visual	Pull to sit
Animate auditory	Defensive reaction
Animate visual/auditory	Activity
Alertness	
Range of State	**Regulation of State**
Peak of excitement	Cuddliness
Rapidity of buildup	Consolability
Irritability	Self-quieting
Lability of state	Hand to mouth
Autonomic Stability	
Tremors	
Startles	
Skin color	

TABLE 7
SUMMARY OF ALTERNATIVE MODELS

Model 2A (Orientation & Motor)

Orientation	Motor
Inanimate visual	Tonus
Inanimate auditory	Motor maturity
Animate visual	Pull to sit
Animate auditory	Defensive reaction
Animate visual/auditory	Activity
Alertness	

Model 2B (Orientation & State Range)

Orientation	State
Inanimate visual	Peak of excitement
Inanimate auditory	Rapidity of buildup
Animate visual	Irritability
Animate auditory	Lability of state
Animate visual/auditory	
Alertness	

Model 3A (Orientation, Motor & Range of State)

Orientation	Motor
Inanimate visual	Tonus
Inanimate auditory	Motor maturity
Animate visual	Pull to sit
Animate auditory	Defensive reaction
Animate visual/auditory	Activity
Alertness	

State
Peak of excitement
Rapidity of buildup
Irritability
Lability of state

TABLE 7
SUMMARY OF ALTERNATIVE MODELS

Model 3B (Orientation, State range, & State regulation)

Orientation	State Range
Inanimate visual	Peak of excitement
Inanimate auditory	Rapidity of buildup
Animate visual	Irritability
Animate auditory	Lability of state
Animate visual/auditory	
Alertness	
State Regulation	
Consolability	
Self quieting	
Hand to mouth	

Model 4
(Orientation, Motor, State Range & State Regulation)

Orientation	Motor
Inanimate visual	Tonus
Inanimate auditory	MOtor maturity
Animate visual	Pull to sit
Animate auditory	Defensive reaction
Animate visual/auditory	Activity
Alertness	
Range of State	**Regulation of State**
Peak of excitement	Cuddliness
Rapidity of buildup	Consolability
Irritability	Self-quieting
Lability of state	Hand to mouth

This model was tested within a multi-sample format which tests for invariance of the hypothesized constructs in the model. The statistics which were produced from the procedure included goodness of fit and mean squared residual indices for each of the samples, and a chi-square statistic for the overall multi-sample analysis.

CHAPTER IV

RESULTS

The results section is subdivided into two major categories: the first documents the multi-sample findings and the second describe the individual sample analyses. The multi-sample findings are specifically related to the study hypotheses stated in the Methodology chapter of the paper. These results are presented in terms of overall chi-square values, chi-square ratios, mean goodness of fit indices and mean root mean square residual findings. The overall chi-square value is evidence for the fit of the factor model to the actual data. The mean fit indices compare the fit of the various models tested.

The individual sample findings are presented in terms of individual sample chi-square, goodness of fit, adjusted goodness of fit, root mean square residual, squared multiple correlations, and maximum likelihood estimates. Each of these values are used to compare models in order that the most optimal model for each sample could be delineated.

Multi-sample Results

Hypothesis 1

The primary goal of this investigation was to establish a construct model of infant behavior based upon NBAS items.

Adequacy of a construct model was substantiated through the investigation of construct invariance across different samples of newborn infants. Invariance was defined in terms of invariance of the number of constructs, parameter estimates, error estimates, and correlation among constructs. Two major models were tested in this investigation: Lester's data reduction system, and a two factor model which included state range and orientation items. It was hypothesized that the Lester model would not be confirmed and that the two factor model would be confirmed with the multi-sample confirmatory factor analyses.

The multi-sample confirmatory factor analysis failed to confirm the hypothesis of equal number of constructs across populations for the Lester (five factors) model. The chi-square value for the multi-sample analysis of the Lester model was found to be 1797 with 597 degree of freedom (df). This value was noted to be significant ($p<.001$), indicating a significant difference between the hypothesized model and the actual data. Given that the number of constructs could not be confirmed, invariance hypotheses concerning parameter estimates, error terms and relationships among the factors could not be tested and were assumed to vary significantly across populations of newborn infants. These findings are considered support for the first hypothesis of this investigation.

Hypothesis 2

The multi-sample analysis of the orientation/state range model (Model 2B) produced a chi-square value of 462 (df=102). This value was noted to be significant ($p<.001$), which indicated a lack of fit for the hypothesized model to the actual data. The lack of confirmation of this two factor model did not support the second hypothesis of this investigation.

Alternative Model Results

The multi-sample results for the alternative models (Models 2A, 3A, 3B, and 4) indicated that none of these models could be confirmed with the multi-sample factor analysis. The chi-square value for Model 2A was 444 (df=129). For Models 3A and 3B the chi-square values were 837 (df=261) and 762 (df=186) respectively. The chi-square value for Model 4 was noted to be 1404 (df=438). In each of these models, the chi-square value was found to be significant ($p < .001$), indicating that these construct models could be consistently defined across the three samples of newborn infants.

Comparison of the Model Fit Indices

In order to further investigate the multi-sample fit of the models, goodness of fit indices (GFI) and root mean square residuals (RMSR) were analyzed. In the multi-sample analysis, GFI and RMSR were calculated for each sample individually. Therefore, mean GFI, RMSR and a chi-square per df ratio were computed to facilitate model comparison. Lower chi-square ratios were indications of better relative fit of the data to the model. Higher mean GFI indices and lower RMSR were considered evidence of better fit for a given model.

Analysis of the chi-square ratios indicated that the smallest value was associated with the Lester model. Other low ratios (3.21) were associated with Models 4 and 3A. The highest mean GFI was associated with Model 2A (mean = .870). The lowest mean GFI was associated with the Lester model (mean = .763). The lowest mean residual was calculated for Model 2A (.169). Other low mean residual values were noted in Models 2B (.189) and 3A (.183). The highest mean residual value was associated with Model 3B (.292). A complete summary of the chi-squares, chi-square ratios, GFI, mean GFI and RMSR values is presented in Table 8.

It should be noted that each of the factor models in the multi-sample analyses had a matrix that was not found to be positive definite. In the Lester model, the phi matrix was not positive definite for the normal sample. In Model 4, the phi matrix was not positive definite for both the drug and normal samples. In Model 3A, the phi matrix in the normal sample and the theta delta matrix in the preterm sample were not positive definite. In Model 3B, the phi matrix was not positive definite for the normal sample. In Model 2A, the theta delta was not positive definite for the preterm sample. For Model 2B, the phi matrix was not positive definite for the normal sample. These matrices violate the assumptions of the

TABLE 8
SUMMARY OF MULTI-SAMPLE MODEL ANALYSES

Lester Model (Five Factor Model)

	GFI	Residual	Chi-square	Chi-square/df
Normal	.689	.351	1797 (597 df)	3.01
Drug	.819	.257		
Preterm	.780	.235		
$\bar{X}$	.763	.281		

Model 4 (orientation, motor, state range & state regulation)

	GFI	Residual	Chi-square	Chi-square/df
Normal	.723	.334	1404 (438 df)	3.21
Drug	.837	.242		
Preterm	.777	.245		
$\bar{X}$	.779	.273		

Model 3A (orientation, motor & state range)

	GFI	Residual	Chi-square	Chi-square/df
Normal	.778	.215	837 (261 df)	3.21
Drug	.869	.178		
Preterm	.831	.157		
$\bar{X}$	.826	.183		

Model 3B (orientation, state range & state regulation)

	GFI	Residual	Chi-square	Chi-square/df
Normal	.805	.322	762 (186 df)	4.09
Drug	.849	.280		
Preterm	.822	.273		
$\bar{X}$	.825	.292		

TABLE 8

TABLE 8 SUMMARY OF MULTI-SAMPLE MODEL ANALYSIS

Model 2A (orientation & Motor)

	GFI	Residual	Chi-square	Chi-square/df
Normal	.810	.228	444 (129 df)	3.44
Drug	.892	.164		
Preterm	.909	.116		
$\bar{X}$	.870	.169		

Model 2B (orientation & state range)

	GFI	Residual	Chi-square	Chi-square/df
Normal	.845	.225	462 (102 df)	4.54
Drug	.877	.185		
Preterm	.860	.157		
$\bar{X}$	.861	.189		

procedures and therefore make the results of the multi-sample analysis suspect.

Individual Sample Analysis

Given that the multi-sample analysis did not confirm any of the models, single sample analyses were undertaken in order to determine a construct model for each individual sample. A summary of the multi-sample model fit indices by sample is presented in Table 9. The models with the highest GFI for each sample were chosen for further confirmatory analysis. For the drug exposed and preterm sample, Model 2A was chosen. For the normal sample, both Model 2A and 2B were chosen because of their equivalent GFI indices. Each of these models were tested in a single sample confirmatory procedure. For model comparison, the Lester model was also tested for each of the individual samples.

Results for each of the individual samples included maximum likelihood estimates for the items, correlation among the factors, squared multiple correlations for individual items, chi-square, GFI, adjusted goodnes of fit (AGFI) and RMSR. Note that GFI and RMSR findings in the individual analyses were equivalent to the the multi-sample procedure.

Normal Sample

The chi-square value for the Lester model model in the normal sample was found to be 627 (df=199 p<.001). The GFI for this model was .689 with an AGFI of .605. The RMSR for this model was found to be .351. The Lester model results

TABLE 9
SUMMARY OF MULTI-SAMPLE FIT INDICES BY SAMPLE

Normal Sample

	GFI	Mean Residual
Lester model	.689	.351
Model 4	.723	.334
Model 3A	.778	.215
Model 3B	.805	.322
Model 2A	.810	.228
Model 2B	.845	.225

Drug Exposed Sample

	GFI	Mean Residual
Lester model	.819	.257
Model 4	.837	.242
Model 3A	.869	.178
Model 3B	.849	.280
Model 2A	.892	.164
Model 2B	.877	.185

Preterm Sample

	GFI	Mean Residual
Lester model	.780	.235
Model 4	.777	.245
Model 3A	.831	.157
Model 3B	.822	.273
Model 2A	.909	.116
Model 2B	.860	.157

were suspect given that the phi matrix was not found to be positive definite.

The chi-square for Model 2A was noted to be 154 (df=43, p<.001). The GFI was found to be .810 with a RMSR of .228. The AGFI was noted to be .709. Model 2A solved without matrix warnings.

The chi-square value for Model 2B was 111 (df=34, p<.001). The GFI was .845 with an AGFI of .749. The RMSR for this model was .225. These results were also suspect because the phi matrix was not positive definite.

Chi-square per degrees of freedom ratios were computed for each model and were found to be equivalent. Therefore, based upon matrix problems and the fit indices, Model 2A was considered the model which most adequately represents the normal sample. A comparison of the fit indices of the three models for the normal sample are presented in Table 10.

Drug Exposed Sample

The chi-square value for the analysis of Lester's model was found to be 656 (df=199, p<.001). The GFI was .819 with a RMSR of .257. The AGFI for this model was .770. The chi-square value for Model 2A was noted to be 199 (df=43, p<.001). The GFI for this model was .892 with a RMSR of .164. The AGFI was .835. The chi-square ratio for these models were equivalent. Therefore, based upon the fit indices, Model 2A appeared to be the most adequate representation of the

TABLE 10
COMPARISON OF LESTER MODEL MODEL 2A AND MODEL 2B
NORMAL SAMPLE ANALYSIS

	Lester	2A	2B
Chi-square	627*	154*	111*
Degrees of freedom	199	43	34
Chi-square/df ratio	3.2	3.5	3.3
Goodness of Fit (GFI)	.689	.810	.845
Adjusted Goodness of Fit	.605	.709	.749
Root Mean Square Residual	.351	.228	.225

* p<.001

underlying structure in the drug exposed sample. A comparison of these models is presented in Table 11.

Preterm Sample

The analysis of Lester's model in the preterm sample revealed a chi-square value of 511 (df=199, p<.001). The GFI was .780 with a RMSR value of .235. The chi-square value for Model 2A was found to be 80 (df=43, p<.001). The theta delta matrix for this model was not positive definite. Analysis of the residuals indicated an overlap between factors for the item motor maturity. Therefore this item was estimated for both factors in a revised solution for Model 2A. This solution had no matrix warnings and therefore was considered as the Model 2A for the preterm group. A comparison of the fit indices of these 2 models are presented in Table 12. Chi-square ratios were found to be equivalent, therefore the modified Model 2A appears to be the more likely model based upon goodness of fit indices.

Model 2A Results

Complete results for Model 2A are presented in Tables 13, 14, and 15. Table 13 presents the maximum likelihood estimates for each of the samples. Note that the maximum likelihood estimates were only significant for all variables in the drug exposed sample. In the normal sample, maximum likelihood estimates were only significant for the orientation factor. For the preterm sample, estimates were significant for orientation, motor maturity and activity items.

TABLE 11
COMPARISON OF LESTER MODEL AND MODEL 2A
DRUG EXPOSED SAMPLE

	Lester	Model 2A
Chi-square	656*	199*
Degrees of freedom	199	43
Chi-square/df ratio	3.3	4.6
Goodness of Fit	.819	.892
Adjusted Goodness of Fit	.770	.835
Root Mean Square Residual	.257	.164

* p < .001

TABLE 12
COMPARISON OF LESTER MODEL AND MODEL 2A
PRETERM SAMPLE

	Lester Model	Model 2A
Chi-square	511*	80*
Degrees of Freedom	199	42
Chi-square/df ratio	2.6	1.9
Goodness of Fit	.780	.920
Adjusted Goodness of Fit	.379	.874
Root Mean Square Residual	.235	.116

* $p < .001$

TABLE 13
MAXIMUM LIKELIHOOD ESTIMATES
INDIVIDUAL SAMPLE COMPARISONS
MODEL 2A

	Normal[b]	Drug Exposed[c]	Preterm[d]
	Orientation Items		
Inanimate Visual	1.000[a]	1.000[a]	1.000[a]
Inanimate Auditory	0.760	0.660	0.171
Animate Visual	1.010	1.040	0.850
Animate Auditory	0.720	0.760	0.420
Animate Visual/Auditory	0.970	1.050	0.850
Alertness	0.940	1.060	0.760
	Motor Items		
Tonus	1.000[a]	1.000[a]	1.000[a]
Motor Maturity	5.830	1.430	0.310
Pull to Sit	1.750	1.020	-0.025
Defensive Reaction	2.960	1.310	0.070
Activity	0.930	0.731	0.270

[a] Estimate fixed to value of 1.000.

[b] Estimates significant for orientation items only ($p < .05$).

[c] Estimates significant for all items ($p < .05$).

[d] Estimates significant for orientation items, Motor Maturity, and Activity ($p < .05$)

Squared multiple correlations for Model 2A are presented in Table 14. In terms of squared multiple correlations, the results showed higher coefficients for orientation items. Motor coefficients were found to be more variable especially within the preterm sample. For example, coefficients were calculated to be less than .1 for the pull to sit, defensive reaction, and the activity items in the preterm sample.

Fit indices for Model 2A are presented in Table 15. Comparison of these indices indicated that the preterm sample data best fit Model 2A. Preterm results indicated a smaller chi-square value (80), a higher GFI (.920), smaller RMSR (.116). The normal sample was associated with the lowest GFI (.810) and the greatest RMSR (.228).

The correlations between the two underlying constructs in Model 2A (orientation and motor functioning) varied among the samples. The highest and only significant correlation between constructs was noted in the drug exposed sample (.778, $p < .05$). The correlation in the normal sample was noted to be .480; the preterm correlation between constructs was the lowest (.081). This low correlation was due, in part, to the motor maturity item being estimated for both constructs.

TABLE 14
SQUARED MULTIPLE CORRELATIONS
INDIVIDUAL SAMPLE COMPARISONS
MODEL 2A

	Normal	Drug Exposed	Preterm
Inanimate Visual	.760	.600	.580
Inanimate Auditory	.450	.400	.070
Animate Visual	.890	.770	.690
Animate Auditory	.390	.500	.170
Animate Visual/Auditory	.800	.730	.440
Alertness	.660	.800	.550
Tonus	.050	.190	.770
Motor Maturity	.690	.370	.420
Pull to Sit	.050	.120	.001
Defensive Reaction	.150	.160	.001
Activity	.040	.180	.101

TABLE 15
FIT INDICES
INDIVIDUAL SAMPLE COMPARISONS
MODEL 2A

	Normal	Drug Exposed	Preterm
Chi-square value	154*	199*	80*
df	43	43	42
Goodness of Fit (GFI)	.810	.892	.920
Adjusted Goodness of Fit (AGFI)	.709	.835	.874
Root Mean Square Residual (RMSR)	.228	.164	.116

* $p < .001$

Summary of the Results

The multi-sample factor analysis did not confirm any of the tested models as stable across different groups of infants. These findings confirmed the hypothesis concerning Lester's model but did not support the hypothesis concerning the orientation/state range model. Analysis of the individual samples revealed that Model 2A (orientation and motor) was the model which best fit each of the samples. Although the chi-square value for each of the sample analyses were significant, the GFI indices for Model 2A were relatively high (.810 to .910), indicating relatively good fit of the model to the actual data in each of the three samples.

Comparison of the individual sample maximum likelihood estimates, squared multiple correlations, and the fit indices for Model 2A revealed the following:

1. In terms of maximum likelihood estimates, orientation items were most consistent across the samples. Motor items tended to vary and be non-significant estimates.
2. A similar pattern was noted for the squared multiple correlation values. Orientation tended to be more consistent across the samples and account for more of the measurement variance when compared to the motor items.
3. Correlation between orientation and motor functioning was only significant in the drug exposed sample.
4. Based upon the fit indices, Model 2A appeared to best fit the preterm sample.

CHAPTER V

DISCUSSION

The discussion of the study findings is presented in terms of explanation of the findings and implications for the examination. Explanations of the findings include the possibility that developmental factors influence the reliability of the NBAS. Implications for the examination include modifications necessary to improve the psychometric properties of the NBAS. The discussion concludes with a summary and a concluding statement of recommendations for further research.

Interpretation of the NBAS Findings

Theoretically, an assessment is devised to measure one or more constructs which are hypothesized to represent the underlying domain of the given behavior. In order for the assessment to be valid, items must consistently and accurately measure the particular hypothetical construct. The results of the multi-sample confirmatory factor analysis indicate that the model of constructs commonly used to represent newborn infant behavior (Lester's clusters) is not consistently defined across different groups of infants. NBAS items do not consistently define orientation, motor, state range, state organization and autonomic functioning as defined by Lester

(1983). This finding supports the study hypothesis concerning Lester's construct model and is an indication that Lester's model is not a valid representation of newborn infant behavior.

The multi-sample analysis failed, however, to uncover another adequate model of underlying structure. The significant chi-square findings for all of the hypothesized models indicates that a stable construct model of infant behavior as measured by the NBAS is lacking. The variability of constructs across samples suggests that interpretation and generalizations concerning infant behavior are questionable with the current set of NBAS items.

The lack of consistent behavioral constructs may be a function of the measurement instrument, developmental patterns inherent to the subgroups of infants, or the statistic used to assess model fit. In terms of the instrument, the lack of model confirmation may be related to the reliability of the NBAS items. The findings of the study indicate that the NBAS items do not perform consistently in different sample situations. This unreliability of measurement could preclude the establishment of a consistent construct model for infant behavior.

However, an argument can be made that the measurement inconsistency is related to the inherent behavioral patterns of the subgroups measured. In other words, the lack of stability of behavioral structure may be due to differences in behavioral abilities among the samples. The findings of construct variability may be documentations of developmental differences among the subgroups of infants. For example, perhaps autonomic stability was not found to be a stable factor because of the specific and possibly different effects of prematurity or drug exposure upon infants. The instability of different NBAS constructs could be interpreted as a behavioral manifestation of a given risk factor.

The multi-sample model findings could also be related to the chi-square statistic itself. The chi-square statistic is based upon sample size and therefore is sensitive when sample sizes are large. Small discrepancies in the model can produce significant values when the sample is large. In the current study, the combined sample size was large. Therefore it is possible that the chi-square values were affected by the sample size. The chi-square values for the models are best considered as a means of comparing models rather than as a statistic itself.

Further study is necessary in order to determine the nature of the construct instability within the NBAS. However, if the variability of underlying constructs is found to be a developmental phenomenon, validation of NBAS constructs will still be necessary. A model of NBAS items must be developed such that consistent constructs are defined before cross-sample comparisons of behaviors can be made.

Both the multiple and the individual sample analyses indicate that a 2 factor model of orientation and motor functioning (Model 2A) is most consistently defined across the three samples. These findings failed to confirm the study hypothesis of an orientation and state range model. Rather, an orientation, motor functioning model best approximate the underlying structure of newborn infant behavior as measured by the NBAS.

The identification of the orientation and motor functioning model as the best fitting model is somewhat surprising given that the majority of exploratory factor studies report an orientation and state related model (Sameroff Krafchuk, & Bakow, 1978; Lester et al. 1976; Osofsky & O'Connor, 1977). It should be noted, however, that this model (2A) and the other hypothesized models in this study were not confirmed with the chi-square statistic. The best model was chosen predominately on the basis of goodness of fit indices. Since the difference between fit indices for the orientation state range (Model 2B) and the orientation motor functioning model (Model 2A) was relatively small and the chi-square per df ratios were comparable for these 2 models, a case could be made that the orientation and state range model is a viable construct model as well.

Additional confirmatory research is necessary to clarify the issue of constructs underlying infant behavior. Further studies should explore different combinations of NBAS items. The findings of this and other studies suggest that the underlying structure of the NBAS involves no more than two major constructs. Multiple factor models do not have the consistency required to be considered valid representations of infant behavior.

Since the multi-sample procedure did not confirm Model 2A, it is assumed that significant variability existed among the subgroups in terms of maximum likelihood estimates and squared multiple correlations. Non-statistical analysis of the consistency and effectiveness of maximum likelihood coefficients for Model 2A indicates that the orientation item coefficients were higher and more consistent across the samples when compared to the motor items. Of note, within the orientation construct, the auditory items tend to show the most variability in terms of likelihood coefficients, suggesting that auditory items are less accurate measurements of the underlying orientation factor.

Motor coefficients tend to be lower and variable across the samples. Analysis of motor construct maximum likelihood estimates reveal that these coefficients are not consistently significant. Furthermore, differences across samples range from a high of 5.52 on the item Motor Maturity to a low of .66 for the item Activity. These findings highlight the relative weakness of the items used to represent motor functioning.

Squared multiple correlations are interpreted as reliability coefficients representing the amount of variance accounted for in the observed item by the underlying construct. For each sample, orientation items had the highest multiple correlations and therefore are considered the more reliable items. Auditory items have the lowest reliability coefficients of the orientation items. Estimates for motor items are generally low across the samples indicating that overall, the motor items are not reliable estimates of motor functioning.

Analysis of the consistency of the reliability coefficients reveals a pattern similar to the maximum likelihood estimates. In general, the orientation items are most consistent across samples, with the auditory items showing the most variability within the orientation construct. Reliability coefficients for the motor items tend to be variable across the samples.

Implications

Item Modification

The findings of the current study indicate that either existing NBAS items need modification or additional items should be developed. One type of modification could involve the scaling of NBAS items. For example, the re-scaling of items required by Lester's system could restrict the variability of the item, thereby limiting its measurement ability. Different types of scaling transformation may need to be explored in order to develop a more adequate recoding of the individual scales. Jacobson et al. (1984), for example, suggest that the NBAS items should not be recoded for analysis. These authors found that clusters were more reliable if they were measured in terms of their original scores.

Other types of modification could involve the criteria used for scaling the NBAS items. Unnecessary variability could be introduced into the NBAS because the scoring criteria involve more than one behavioral dimension. For example, the criteria for scoring alertness involves both amount and quality of alertness. Scaling of this item might be improved if the scale included only one aspect of alertness: quality of alertness and amount of alertness.

This rescaling concept can be applied to other NBAS items, such as general tone. This item assesses the tone of the infant but does not distinguish between trunk tone and extremity tone. It is conceivable that an infant can have a hypotonic trunk and hypertonic extremities. The scale would become more consistent if the item was separated to distinguish between trunk and extremities tone. In general, more specific definition of items could serve to improve the reliability and validity of the NBAS.

Construct Modification

In light of the findings of this investigation, reconceptualization of hypothetical constructs may also be considered for the NBAS. Underlying constructs such as state regulation or state range may be to global to reliably measure. Items such as irritability, and lability of state might be concepts from which new constructs could be developed.

For example, "irritability" items could be developed which measure the amount of facilitation required to calm an infant. This can be added to the current consolability item which documents the type of manuever required to calm an infant after he or she has become upset (Brazelton, 1984). Another aspect of irritability could be similar to peak of excitement; that is, what type of manuevers tend to upset the infant? These types of item additions and reorganizations could provide the stable constructs with which comparative analysis with the NBAS could occur.

Summary of Modifications

To summarize the possible modifications of the NBAS, it is apparent that the orientation dimension is the most consistent and stable construct measured by the NBAS items. Additional item development for this construct may need to center on the auditory items as they have the weakest measurement qualities. The motor construct is much less stable than the orientation construct. Therefore, additional motor items and modification of the scales of existing motor items are necessary in order that motor functioning be adequately measured. Autonomic functioning, state range, and state organization are not definable as constructs with the current set of NBAS items. Items related to these concepts need to be redefined and re-scaled in order for them to be considered as possible dimensions of newborn infant behavior.

Usage of the NBAS

The findings of the current study bring into question the validity of the NBAS as a research tool. Quantitative NBAS information concerning underlying constructs are not identified as reliable representations of newborn infant behaviors. In its current state, the NBAS cannot be used to compare different groups of infants in a valid manner.

These negative results, however, do not diminish the clinical importance of the examination. The current items on the examination provide the clinician with critical information concerning the current behavior of the infant. This type of information is valuable for clinical diagnosis and creations of interventions beginning in the newborn period. The NBAS can be especially helpful to parents as they can learn information concerning early social behavior of their infant.

The goals of clinical relevance and psychometric integrity may be mutually exclusive in the newborn period. It is conceivable that the inherently variable nature of infancy requires a flexible assessment to be developmentally relevant. Given that it appears that only minimal modification of the NBAS has been attempted since its publication, this type of rationalization seems premature. Further investigations into modifications of the NBAS are necessary before infants are considered too variable to examine validly. Thoughtful expansion and reconceptualization of the examination could greatly improve the measurement qualities of the examination without reducing its clinical sensitivity.

This is the direction that continued research with the NBAS must take. Ongoing clarification of the abilities of the newborn infant are dependent upon the integrity of the measurement instrument. A valid assessment of newborn behavior could improve the understanding of relationships between early behaviors and later developmental outcomes. This type of knowledge base would improve the clinician's ability to validly diagnose at-risk infants and provide the family with the appropriate developmental support needed to ameliorate possible developmental dysfunctions.

Summary and Conclusions

The multi-sample confirmatory factor analysis failed to confirm Lester's model of underlying constructs. The hypothesized two factor model of orientation and state range and other alternative models were also not confirmed by the multi-sample procedure. Overall, these results indicate that items in the NBAS react differently in different samples of infants. The variability of NBAS items could be a function of the examination or the high risk nature of the samples. In either case, validity of constructs commonly used to represent infant behavior was not established.

Individual sample analyses revealed that a model of orientation and motor constructs is the model which best fit each of the tested samples. However, similarity of fit findings indicate that other alternative construct models for NBAS data could be considered. Significant variablity in the definition of the constructs is obvious in the analyses. Therefore, it cannot be said that any of the NBAS constructs, as they are currently defined, are reliable and valid.

The findings of the current study raise doubts concerning the translation of NBAS scales into quantifiable representations of infant behavior. Continued research is necessary in order that the nature of underlying behavioral constructs in infancy be documented. These types of issues must be clarified before the NBAS can be considered a valid measurement of infant behavior. Further research with the NBAS should include the following areas:

1. Investigation of different scaling methods for NBAS items.
2. Development of additional NBAS items.
3. Redefintion of hypothetical constructs or the development of new constructs.
4. Confirmatory factor analytic studies of the modified NBAS.

BIBLIOGRAPHY

Adamson, L., Als, H., Tronick, E., & Brazelton, T. (1975). The development of social reciprocity between a sighted infant and her blind parent. Journal of the American Academy of Child Psychiatry, 16, 194-207.

Aleksandrowicz, M. & Aleksandrowicz, D. (1976). Precursors of ego in neonates: Factor analysis of Brazelton scale data. Journal of the American Academy of Child Psychiatry, 15, 257-268.

Als, H. (1978). Assessing an assessment: Conceptual considerations, methodological issues, and a perspective on the future of the Neonatal Behavioral Assessment Scale. In A. Sameroff (Ed.), Organization and stability of newborn behavior: A commentary on the Brazelton Neonatal Behavioral Assessment Scale. Monographs of the Society for Research in Child Development, 43, No 5-6.

Als, H. (1983). Infant individuality: Assessing patterns of very early development. In J. Galenson & E. Tyson (Eds.), Frontiers of infant psychiatry, (pp. 363-378). New York: Basic Books.

Als, H. (1986). A synactive model of neonatal behavioral organization: Framework for the assessment of neurobehavioral development in the premature infant and support of infants and parents in the neonatal intensive care environment. Physical and Occupational Therapy in Pediatrics, 6, 3-52.

Als, H., Lester, B., Tronick, E., & Brazelton, T. (1982). Manual for the assessment of preterm infants' behavior (APIB). In H. Fitzgerald, B. Lester, & M Yogman (Eds.), Theory and research in behavioral pediatrics, (Vol. 1) (pp. 35-132). New York: Plenum.

Als, H., Tronick, E., Lester, B., & Brazelton, T. (1977). The Brazelton Neonatal Behavioral Assessment Scale (BNBAS). Journal of Abnormal Child Psychology, 5, 215-237.

Als, H., Tronick, E., Lester, B., & Brazelton, T. (1979). The Brazelton Neonatal Scale (BNBAS). In J. Osofsky (Ed.), Handbook of Infant Development (pp. 185-215). New York: Wiley.

Anastasi A. (1982). Psychological Testing (5th ed.). New York: Macmillian.

Angoff, W. (1988). Validity: An evolving concept. In H. Wainer & H. Braun (Eds.), Test Validity (pp. 19-33). New Jersey: Erlbaum.

Bagozzi R. (1978). The construct validity of the affective, behavioral and cognitive components of attitued by analysis of covariance structures. Multivariate Behavioral Research, 13, 9-31.

Bagozzi, R. (1981). an examination of the validity of two models of attitude. Multivariate Behavioral Research, 16, 323-359.

Bentler, P. (1980). Multivariate analysis with latent variables. Causal models. Annual Review of Psychology, 31, 419-456.

Bentler, P., & Weeks, D. (1980). Linear structural equations with latent variables. Psychometrika, 45, 289-308.

Biddle, B., & Marlin, M. (1987). Causality, confirmation, credulity, and structural equation modeling. Child Development, 58, 4-17.

Brazelton, T. (1973). Neonatal Behavioral Assessment Scale. Clinics in Developmental Medicine (No. 50). London: Heinemann.

Brazelton, T. (1984). Neonatal Behavioral Assessment Scale 2nd ed. Spastics International Medical Publications. London: Blackwell.

Brazelton, T., Nugent, J., & Lester, B. (1987). Neonatal Behavioral Assessment Scale. In J. Osofsky (Ed.), Handbook of infant development (pp. 780-817). New York: Wiley.

Brazelton, T., Tryphonopoulou, Y., & Lester, B. (1979). A comparative study of the behavior of Greek neonates. Pediatrics, 63, 279-285.

Carroll, J. (1953). An analytic solution for approximating simple structure in factor analysis. Psychometrika, 18, 23-38.

Carroll, J. (1957). Biquartimin criterion for rotation to oblique simple structure in factor analysis. Science, 126, 1114-1115.

Chasnoff, I., Griffith, D., MacGregor, S., Dirkes, K., & Burns, K. (1989). Temporal patterns of cocaine use in pregnancy: Perinatal outcome. Journal of the American Medical Association, 261, 1741-1744.

Chasnoff, I., Hatcher, R., & Burns, W. (1982). Polydrug-and methadone-addicted newborns: A continum of impairment? Pediatrics, 70, 210-213.

Crockenberg, S. (1981). Infant irritability, mother responsiveness and social support influences on the security of infant-mother attachment. Child Development, 52, 857-865.

Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16, 297-334.

Darwin, C. (1877). A biographical sketch of an infant. Mind, 2, 286-294.

Dixon, S., Keefer, C., Tronick, E., & Brazelton, T. (1982). Perinatal circumstances and neborn outcome among the Gusii of Kenya: Assessment of risk. Infant Behavior and Development, 5, 11-32.

Entwisle, D., & Hayduk, L. (1981). Academic expectations and the school attainment of young children. Sociology of Education, 54, 34-50.

Ethington, C., & Wolfle, L. (1986). A structural model of mathematics achievement for men and women. American Educational Research Journal, 23, 65-75.

Ferrari, F., Grosoli, M., Fontane, G., & Cavazutti, G. (1983). Neurobehavioral comparison of low-risk preterm and fullterm infants at term conceptual age. Developmental Medicine and Child Neurology, 25, 450-458.

Fleishman, J., & Benson, J. (1987). Using LISREL to evaluate measurement models and scale reliablilty. Educational and Psychological Measurement, 47, 925-939.

Francis, P., Self, P., & Horowitz, F. (1987). The behavioral assessment of the neonate: An overview. In J. Osofsky (Ed.), Handbook of infant development (pp 723-779). New York: Wiley.

Garcia-Coll, C., Sepkoski, C., & Lester, B. (1981). Cultural and biomedical correlates of neonatal behavior. Developmental Pediatrics, 14, 147-154.

Garcia-Coll, C., Sepkoski, C., & Lester, B. (1982). Effects of teenage childbearing on neonatal and infant behavior in Puerto Rico. Infant Behavior and Development, 5, 227-236.

Häggulund G. (1982). Factor analysis by instrumental variables. Psychometrika, 47, 209-222.

Harmon, H. (1967). Modern factor analysis. Chicago: University of Chicago Press.

Hay, D. (1986). Infancy. Annual Review of Psychology, 37, 135-161.

Holmes, D., Nagy, J., Slaymaker, F., Sosnowski, R., Prince, S., & Pasternak, J. (1982). Early influences of prematurity, illness, and prolonged hospitalization on infant behavior. Developmental Psychology, 18, 744-750.

Holzinger, K. (1944). A simple method of factor analysis. Psychometrika, 9, 257-262.

Holzinger, K. (1949). Applications of the simple method of factor analysis. Journal of Educational Psychology, 40, 129-142.

Holzinger, K., & Swinefield, F. (1937). The bi-factor method. Psychometrika, 2, 41-54.

Horowitz, F., & Linn, P. (1984). Use of the NBAS in research. In T. B. Brazelton (Ed.), Neonatal Behavioral Assessment Scale (2nd ed). pp. 97-104. Spastics International Medical Publications. London: Blackwell.

Horowitz, F., Sullivan, J., & Linn P., (1978). Stability and instability in the newborn infant: The quest for elusive threads. In A. Sameroff (Ed.), Organization and stability of newborn behavior: A commentary on the Brazelton Neonatal Behavioral Assessment Scale. Monographs of the Society for Research in Child Development, 43, (5-6 Serial No. 177).

Huba, G., Winegard, J., & Bentler, P. (1981). A comparison of two latent variable causal models for adolescent drug use. Journal of Personality and Social Psychology, 40, 180-193.

Jacobson, J., Fein, G., Jacobson, S., & Schwartz, P. (1985). Factors and clusters for the Brazelton scale: An investigation of the dimensions of neonatal behavior. Developmental Psychology, 20, 339-353.

Jenrich, R., & Sampson, P. (1966). Rotation for simple loadings. Psychometrika, 31, 165-178.

Jöreskog, K. (1967). Some contributions to maximum likelihood factor analysis. Psychometrika, 31, 443-482.

Jöreskog, K. (1969). A general approach to confirmatory maximum likelihood factor analysis. Psychometrika, 34, 183-202.

Jöreskog, K. (1983). Factor analysis as an errors-in-variables model. In H. Wainer & S. Messick (Eds.), Principles on Modern Psychological Measurement: A Festschrift in Honor of Frederic M. Lord (pp.185-196). Hillsdale, N.J.: Erlbaum.

Jöreskog, K., & Sörbom, D. (1988). LISREL 7: A guide to the program and applications (2nd ed.) Chicago: SPSS Inc.

Kloek, T., & Mennes, L. (1960). Simultaneous equations estimation based on principle components of predetermined variables. Econometrica, 28, 45-61.

Kuhnert, B., Harrison, M., Linn, P., & Kuhnert, P. (1984). Effects of maternal epidural anesthesia on neonatal behavior. Anesthesia & Analgesia, 63, 301-308.

Lamb, M., & Bornstein, M. (1987). Development in infancy: An introduction (2nd ed.). New York: Random House.

Lester, B. (1983). A method for study of change in neonatal behavior. In T. Brazelton & B. Lester (Eds.), Infants at risk: Assessment and intervention. New York: Elsevier.

Lester, B. (1984). Data analysis and prediction. In T.B. Brazelton (Ed.), Neonatal Behavioral Assessment Scale (2nd ed.) (pp. 85-96). Spastics International Medical Publications. London: Blackwell.

Lester, B., Als, H., & Brazelton, B. (1982). Regional obstetric anesthesia and newborn behavior: A reanalysis toward synergistic effects. Child Development, 53, 687-692.

Lester, B., Emory, E., Hoffman, S., & Eitzman, D. (1976). A multivariate study of the effects of high risk factors on performance on the NBAS. Child Development, 47, 515-517.

Macrae, D. (1960). Direct factor analysis of sociometric data. Sociometry, 23, 360-370.

McArdle, J., & McDonald, R. (1984). Some algebraic properties of the reticular action model for moment structures. British Journal of Mathematical Statistical Psychology, 37, 234-251.

Kaiser, H. (1958). The varimax criterion for analytic rotation in factor analysis. Educational and Psychological Measurement, 23, 770-773.

Kaye K. (1978). Discriminating among normal infants by multivariate analysis of Brazelton scores: Lumping and smoothing. In A. Sameroff (Ed.), Organization and stability of newborn behavior: A commentary on the Brazelton Neonatal Behavioral Assessment Scale. Monographs of the Society for Research in Child Development, 43, (5-6 Serial No. 177).

Kesterman, G. (1981). Assessment of individual differences among healthy newborns on the Brazelton Scale. Early Human Development, 5, 15-27.

Kim, J., & Mueller, C. (1978). Factor Analysis: Statistical methods and practical issues. Sage University Paper Series on Quantitative Application in the Social Sciences, 07-014. Beverly Hills and London: Sage Publications.

Marsh, H., & Hocevar, D. (1983). Confirmatory factor analysis of multi-trait, multi-method matrices. Journal of Educational Measurement, 20, 231-248.

√Maruyama, G., Rubin, R., & Kingsbury, G. (1981). Self-esteem and educational achievement: Independent constructs with a common cause. Journal of Personality and Social Psychology, 40, 962-975.

McDonald, R. (1978). A simple comprehensive model for the analysis of covariance structures. British Journal of Mathematical and Statistical Psychology, 31, 59-72.

McGaw, B., & Jöreskog, K. (1971). Factorial invariance of ability measures in groups differing in intelligence and socio-economic status. British Journal of Mathematical and Statistical Psychology, 24, 154-168.

Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. In H. Wainer & H. Braun (Eds.), Test Validity, (pp. 33-47), New Jersey: Erlbaum.

Mulaik, S. (1986). Factor analysis and Psychometrika: Major developments. Psychometrika, 51, 23-33.

Nugent, J., & Sepkoski C. (1984). The training of NBAS examiners. In T. Brazelton (Ed.), Neonatal Behavioral Assessment Scale (2nd ed) (pp. 78-84). Spastics International Medical Publications. London: Blackwell.

Newton, R., Komaeoka, V., Hoelter, J., & Tanaka-Matsumi, J. (1984). Maximum likelihood estimation of factor structures of anxiety measures: A multiple group comparison. Educational and Psychological Measurement, 44,179-193.

Nunnally, J. (1978). Psychometric Theory (2nd ed.). New York: McGraw-Hill.

Osofsky, J., & Danzger, B. (1974). Relationships between neonatal characteristics and mother-infant interaction. Developmental Psychology, 10, 124-130.

Osfosky, J., & O'Connell, E. (1977). Patterning of newborn behavior in an urban population. Child Development, 48, 532-536.

Overall, J., & Hollister, L. (1964). Computer procedures for psychiatric classification. Journal of the American Medical Association, 187, 64-83.

Paludetto R., Mansi, G., Rinaldi, P., DeLuca, T., Corchia, C., DeCurtis, M., & Andolfi, M. (1982). Behavior of preterm newborns reaching term without any serious disorder. Early Human Development, 6, 357-363.

Papile, L., Munsick-Bruno, G., & Schaefer, A. (1983). Relationship of cerebral intraventricular hemorrhage and early childhood neurologic handicaps. Journal of Pediatrics, 103, 273-277.

Petrinow, L., & Hardyck, C. (1964). Behavioral changes in Parkinson patients following surgery-Factor analytic study. Journal of Chronic Diseases, 17, 13-24.

Prechtel, H., & Beintema, J. (1968). The neurological examination of the full-term newborn infant. Clinics in developmental medicine No. 28. London: Spastics International Medical Publications with Heinemann Medical.

Sameroff, A., & Chandler, M. (1975). Reproductive risk and the continuum of caretaking casualty. In F. Horowitz, M. Hetherington, S. Scarr-Salapatek, & G. Siegel (Eds.), Review of child development research (Vol. 4), (pp.187-244). Chicago: University of Chicago Press.

Sameroff, A., Krafchuk, E., & Bakow, H., (1978). Issues in grouping items from the neonatal behavioral assessment scale. In A. Sameroff (Ed.), Organization and stability of newborn behavior: A commentary on the Brazelton Neonatal Behavioral Assessment Scale. Mongraphs of the Society for Research in Child Development, 43, (5-6 Serial No 177).

Saunders, D. (1953). An analytic method for rotation to orthogonal simple structure. Research Bulletin, 53, 10.

Scanlon, K., Scanlon, J., & Tronick, E. (1984). The impact of perinatal and neonatal events on the early behavior of the extremely premature human. *Developmental & Behavioral Pediatrics*, *5*, 65-73.

Scarr-Salapatek, S., & Williams, M. (1973). The effects of early stimulation on low-birthweight infants. *Child Development*, *44*, 94-101.

Sell, E., Luick, A., Poisson, S., & Hill, S. (1980). Outcome of very low birthweight (VLBW) infants. I. Neonatal behavior of 188 infants. *Journal of Developmental and Behavioral Pediatrics*, *1*, 78-85.

Sörbom, D. (1974). a general model for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology*, *27*, 229-239.

Sostek A., & Anders, T. (1977). Relationships among the Brazelton Neonatal Scale, Bayley Infant Scales, and early temperment. *Child Development*, *48*, 320-323.

Spearman, C. (1904). General intelligence objectively determined and measured. *American Journal of Psychology*, *15*, 201-293.

Spearman, C. (1927). *The abilities of man*. New York: Macmillian.

Statistical Package for the Social Sciences. (1983). Chicago: SPSS Incorporated.

Strauss, M., & Rourke, D. (1978). A multivariate analysis of the Neonatal Behavioral Assessment Scale in several samples. In A. Sameroff (Ed.), Organization and stability of newborn behavior: A commentary on the Brazelton Neonatal Behavioral Assessment Scale. Monographs of the Society for Research in Child Development, 43 5-6.

Strauss, M., Lesson-Firestone, J., Starr, R., & Ostrea, E. (1976). Behavior of narcotics addicted newborns. Child Development, 46, 887-893.

Streissguth, A., Barr, H., & Martin, D. (1983). Maternal alcohol use and neonatal habituation assessed with the Brazelton scale. Child Development, 54, 1109-1118.

Tanaka, J., & Huba, G. (1984). Confirmatory hierarchical factor analyses of psychological distress measures. Journal of Personality and Social Psychology, 46, 621-635.

Tatsouka, M. (1988). Multivariate Analysis: Techniques for educational and psychological research (2nd ed.). New York: Macmillian Publishing.

Thurstone, L. (1931). Multiple factor analysis. Psychological Review, 38, 406-427.

Thurstone, L. (1944). Second-order factors. Psychometrika, 9, 71-100.

Thurstone, L. (1945). A multiple group method of factoring the correlation matrix. Psychometrika, 10, 73-78.

Waters, D., Vaughn, B., & Egelund, B. (1980). Individual differences in infant-mother attachment relationships at age one: Antecedents in neonatal behavior in an urban, economically disadvantaged sample. Child Development, 51, 208-216.

Thurstone, L. (1947). Multiple factor analysis. Chicago: University of Chicago Press.

Werts, C., Rock, D., Linn, R., & Jöreskog, K. (1977). Validating psychometric assumptions within and between populations. Educational and Psychological Measurement, 37, 863-871.

Woodbury, M., Clelland, R., & Hickey, R. (1963). Applications of a factor analytic model in the prediction of biological data. Behavioral Science, 8, 347-354.

Woodson, R., DaCosta-Woodson, E. (1980). Covariates of analgesia in a clinical sample and their effect on the relationship between analgesia and infant behavior. Infant Behavior and Development, 3, 205-213.

Yogman, M., Cole, P., Als, H., & Lester, B. (1982). Behavior of newborns of diabetic mothers. Infant Behavior and Development, 4, 297-306.

CURRICULUM VITAE

Scott Daniel Azuma was born on September 30, 1957, in Chicago Illinois. Scott graduated from Lane Technical High School in Chicago in 1975. He graduated with a Bachelors of Science in Physical Therapy from the University of Illinois at Chicago in 1979. In 1987, he obtained a Masters of Science in Physical Therapy from Northwestern University. He has been licensed to practice physical therapy in Illinois since 1979.

Scott has worked as a pediatric physical therapist at Childrens Memorial Hospital, Micheal Reese Hospital and Northwestern Memorial Hospital. His therapy roles have included direct treatment, consulation and evaluation. He has treated a wide variety of diagnoses including cerebral palsy, hemophiliacs and premature infants.

APPROVAL SHEET

The dissertation submitted by Scott Azuma has been read and approved by the following committee:

Dr. Jack Kavanagh, Co-director
Professor, Counseling and Educational Psychology, Loyola

Dr. Frances Stott, Co-director
Associate Professor, Counseling and Educational Psychology, Loyola,
Professor, Erikson Institute

Dr. Dan Griffith
National Association for Perinatal Addiction Research and Education

Dr. Robert Halpern
Associate Professor, Counseling and Educational Psychology, Loyola,
Professor, Erikson Institute

The final copies have been examined by the co-directors of the dissertation and the signatures which appears below verify the fact thatt any necessary changes have been incorporated and that the dissertation is now given final approval by the Committee with reference to content and form.

The dissertation is therefore accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

4/17/91
Date

Frances M. Stott
Jack A. Kavanagh
Co-directors Signatures